# PNAS

## www.pnas.org

Supplementary Information for

**An ancient, conserved gene regulatory network led to the rise of oral venom systems**

Agneesh Barua, Alexander S. Mikheyev

Agneesh Barua
Email: agneesh.barua@oist.jp

**This PDF file includes:**

Supplementary Figures 1-6

**Other supplementary materials for this manuscript include the following:**

Supplementary Tables 1 to 5 Supplementary Tables M1 and M2

**Supplementary Information Text**

**Subhead.** Supplementary code, data, figures, and table can be found at https://github.com/agneeshbarua/Metavenom

# Supplementary Materials

Supplementary code, data, figures, and table can be found at

https://github.com/agneeshbarua/Metavenom

We recommend the reader to refer to the above online supplementary material as in allows

for greater flexibility thereby greatly improving its readability and making the review
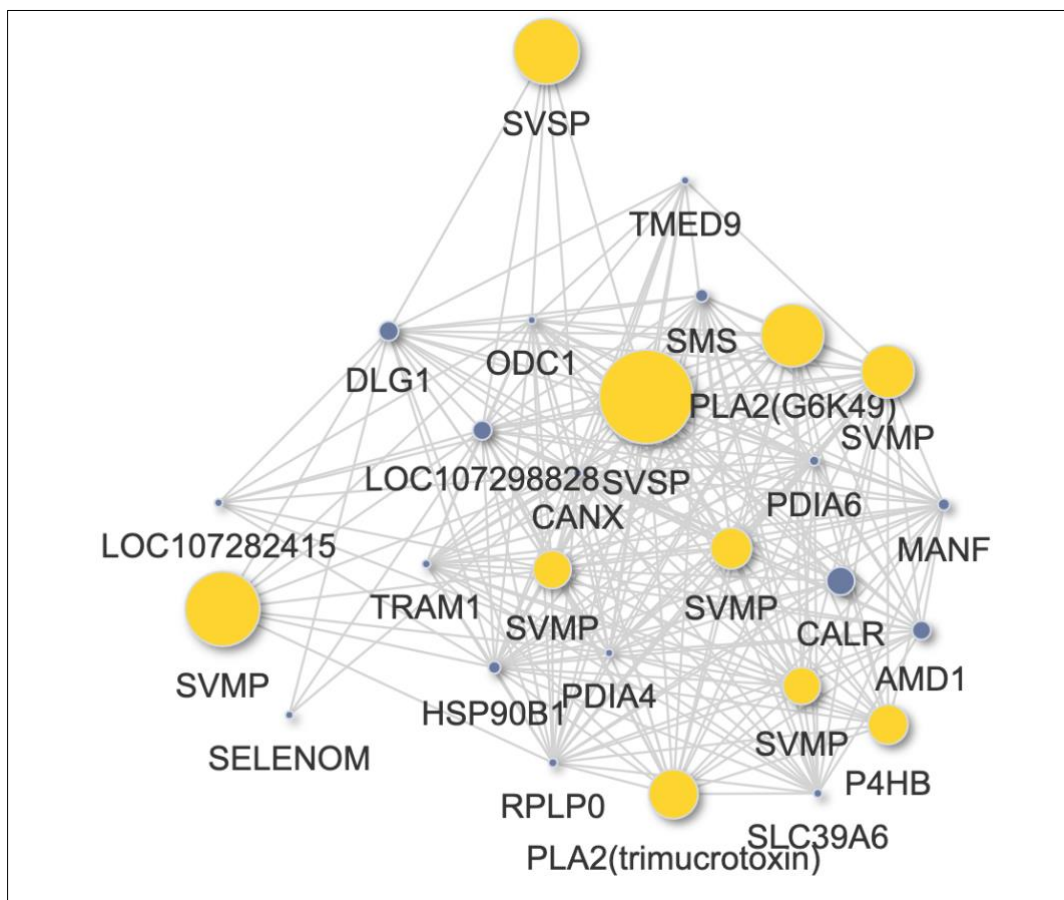
process much easier.

Figure S1: An interactive version of this plot can be found at https://github.com/agneeshbarua/Metavenom. The interactive version allows for better visualization by choosing which type of genes (housekeeping or venom) to display.
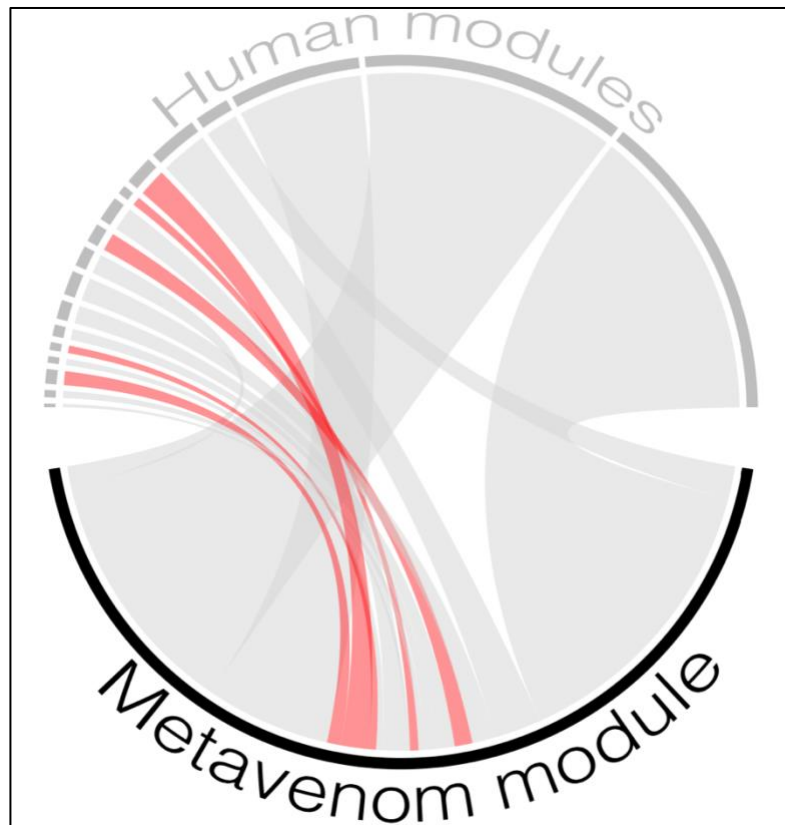


Figure S2: Cross tabulation of meta-venom module was carried out using an external dataset of human salivary gland *(25)*. The red ribbons indicate genes in the meta-venom that overlap with modules in the human salivary gland dataset. We tested for overlap using Fisher's exact tests correcting for multiple comparisons using the Benjamini-Hochberg procedure with the false discovery rate set at 0.05.
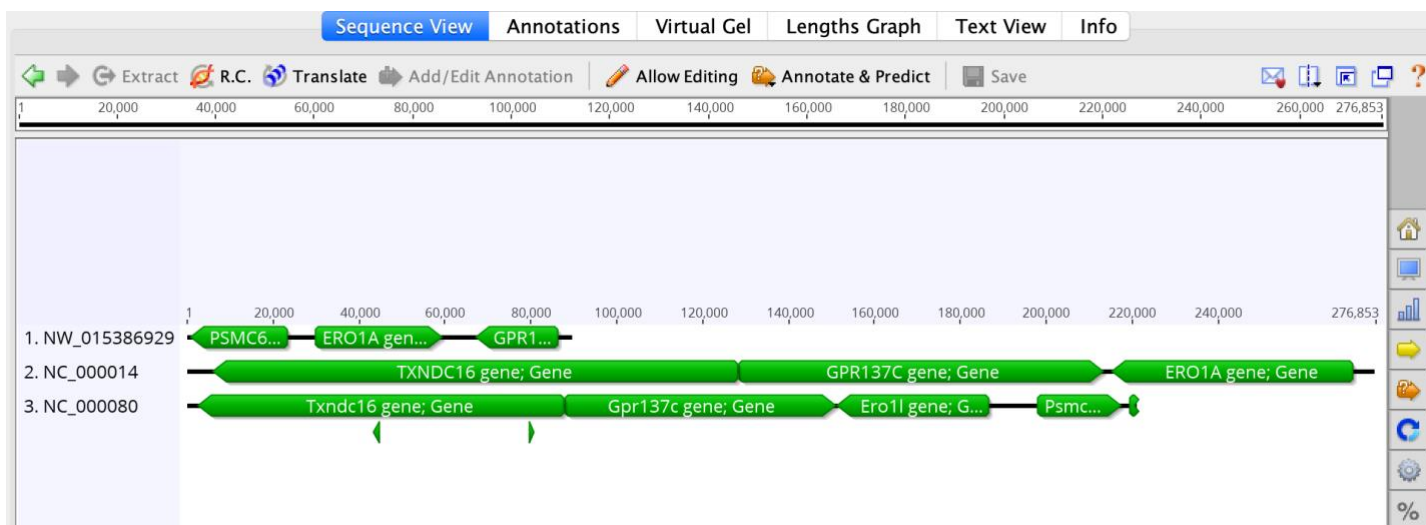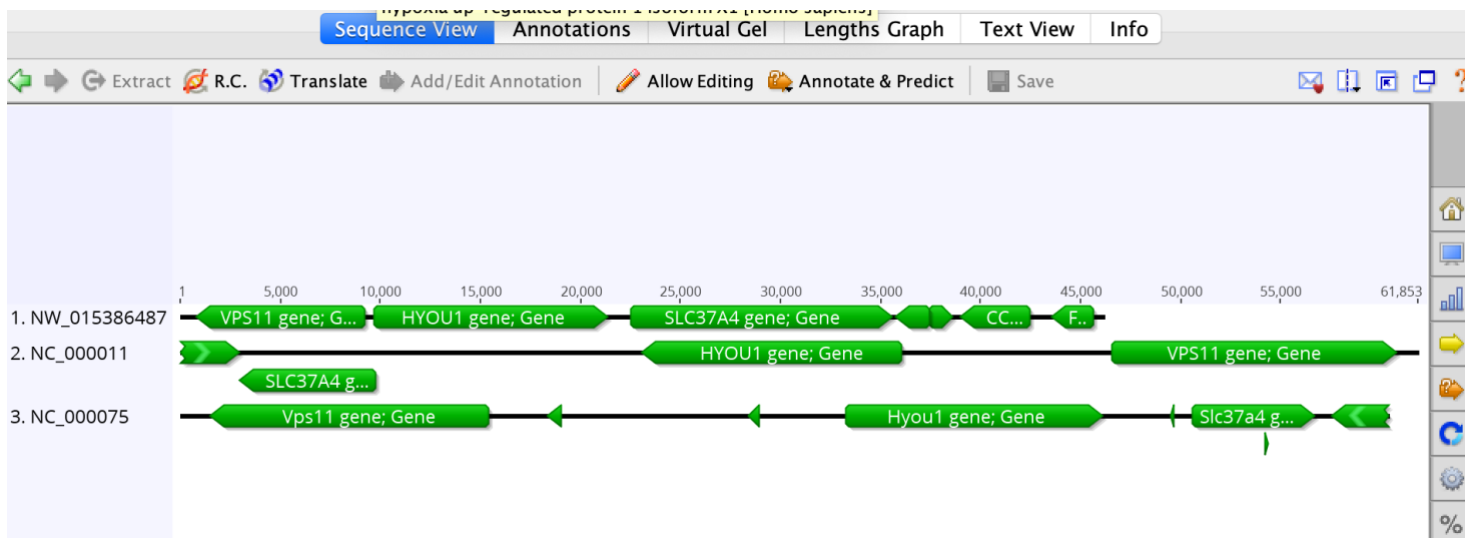
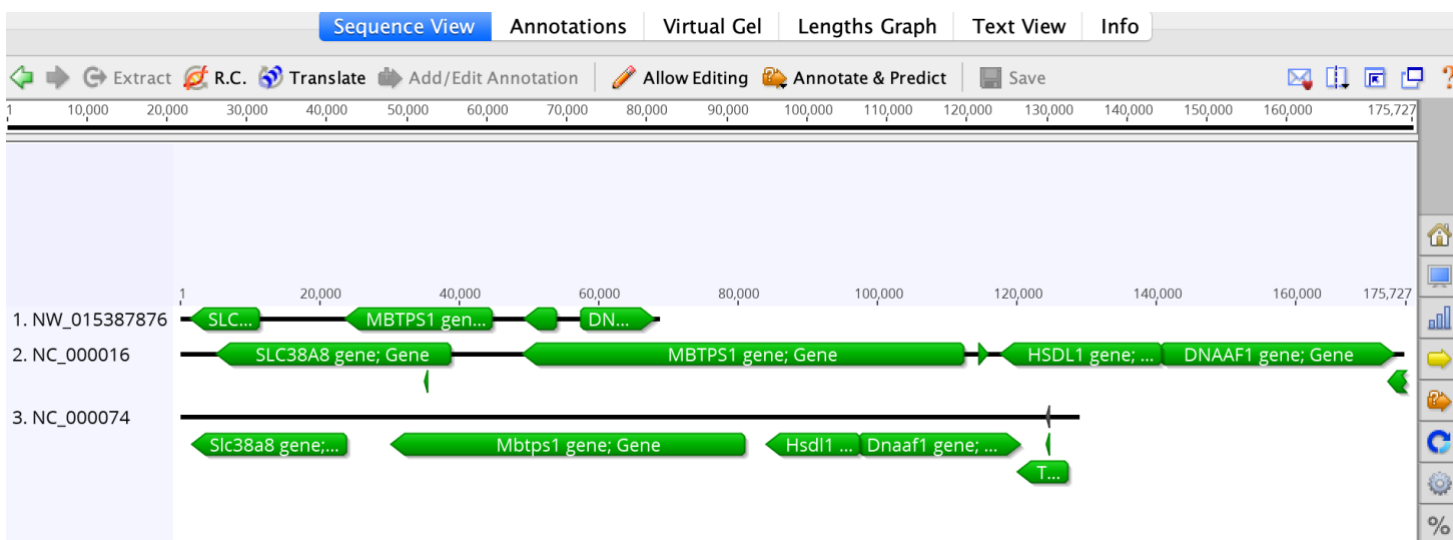**A**
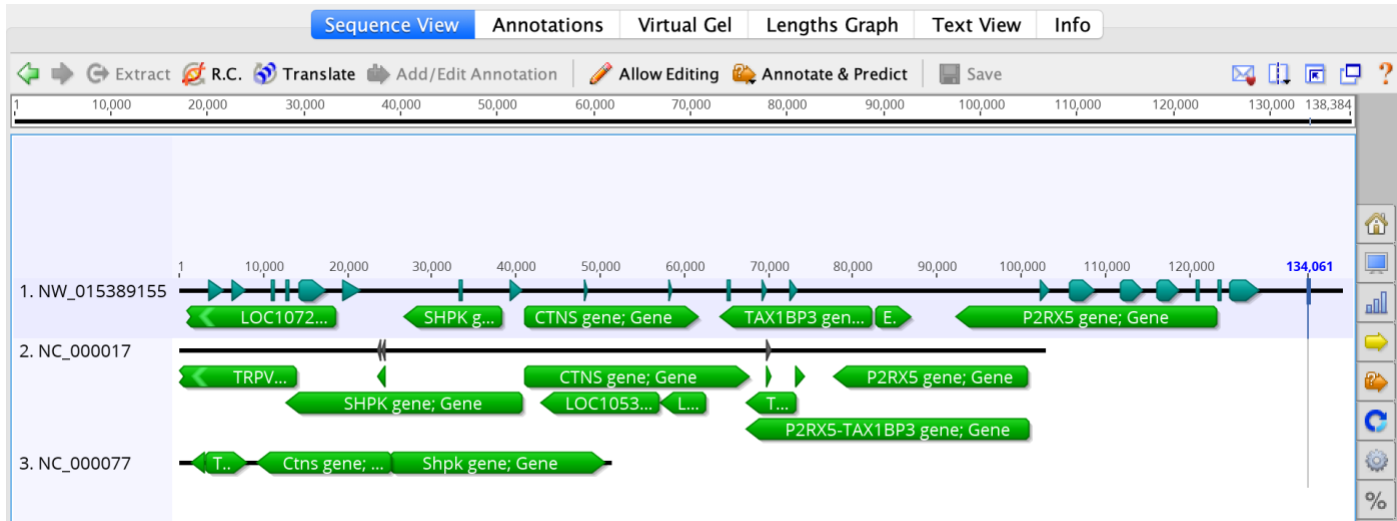


**B**
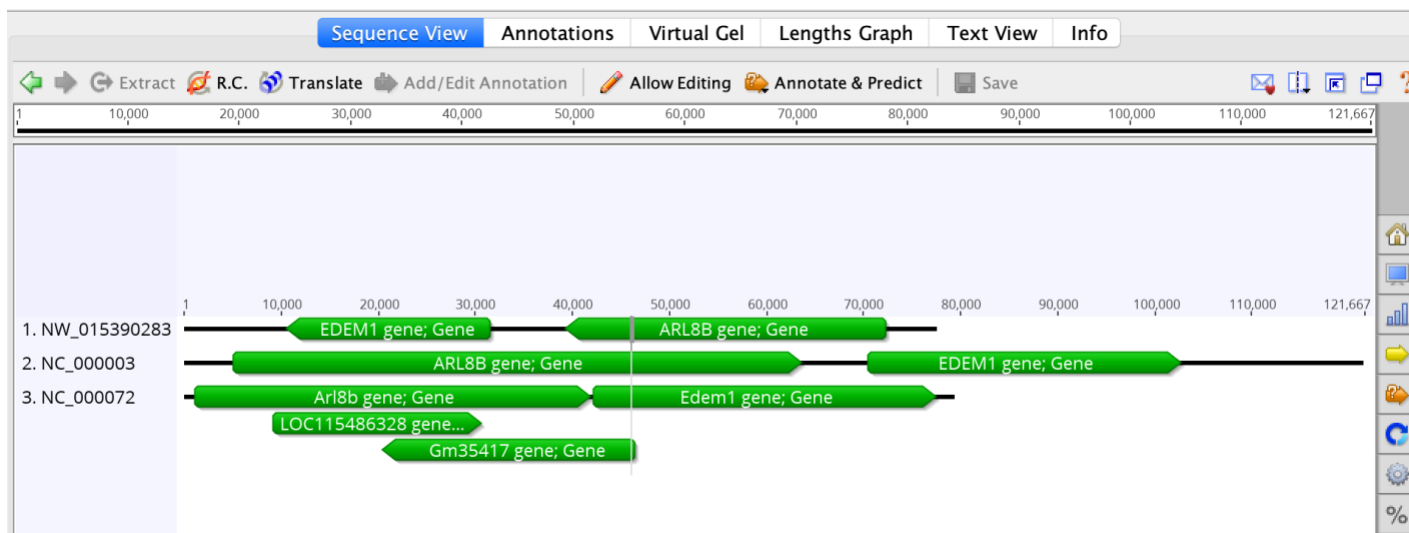
**C**



**D**



**E**

Figure S3: The orthologs estimated by combining NCBI orthologs and OrthoFinder were highly syntenic. The above figures were displayed in Geneious 10.2.6 (https://www.geneious.com). Only 5 gene are displayed. Online supplementary material has information regarding how to analyse more genes. (A) CTNS; GeneID:107297140 (*Protborthrops mucrosquamatus*, Pm), GeneID:1497 (*Homo sapiens*, Hs), GeneID:83429 (*Mus musculus*, Mm). (B) EDEM1; GeneID:107298918 (Pm), GeneID:9695 (Hs), GeneID:192193 (Mm). (C) HYOU1 GeneID:107284796 (Pm), GeneID:10525 (Hs), GeneID:12282 (Ms). (D) ERO1A GeneID:107289096 (Pm), GeneID:30001 (Hs), GeneID:50527 (Mm). (E) MBTPS GeneID:107293778 (Pm), GeneID:8720 (Hs), GeneID:56453 (Mm).
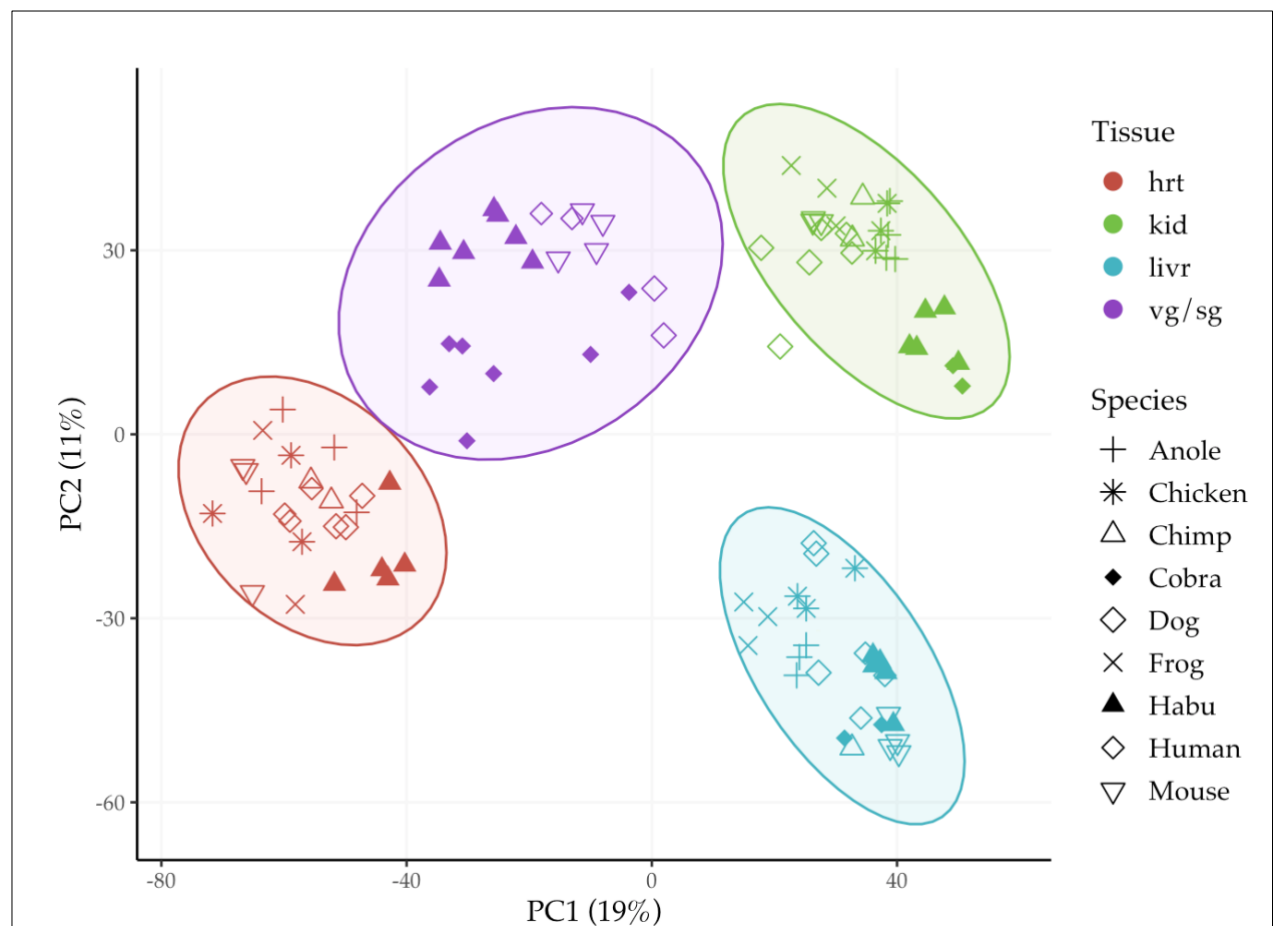


Figure S4: PCA using all 2682 expressed orthologs between 9 taxa, including those outside the meta-venom homologous tissues clustered more tightly.
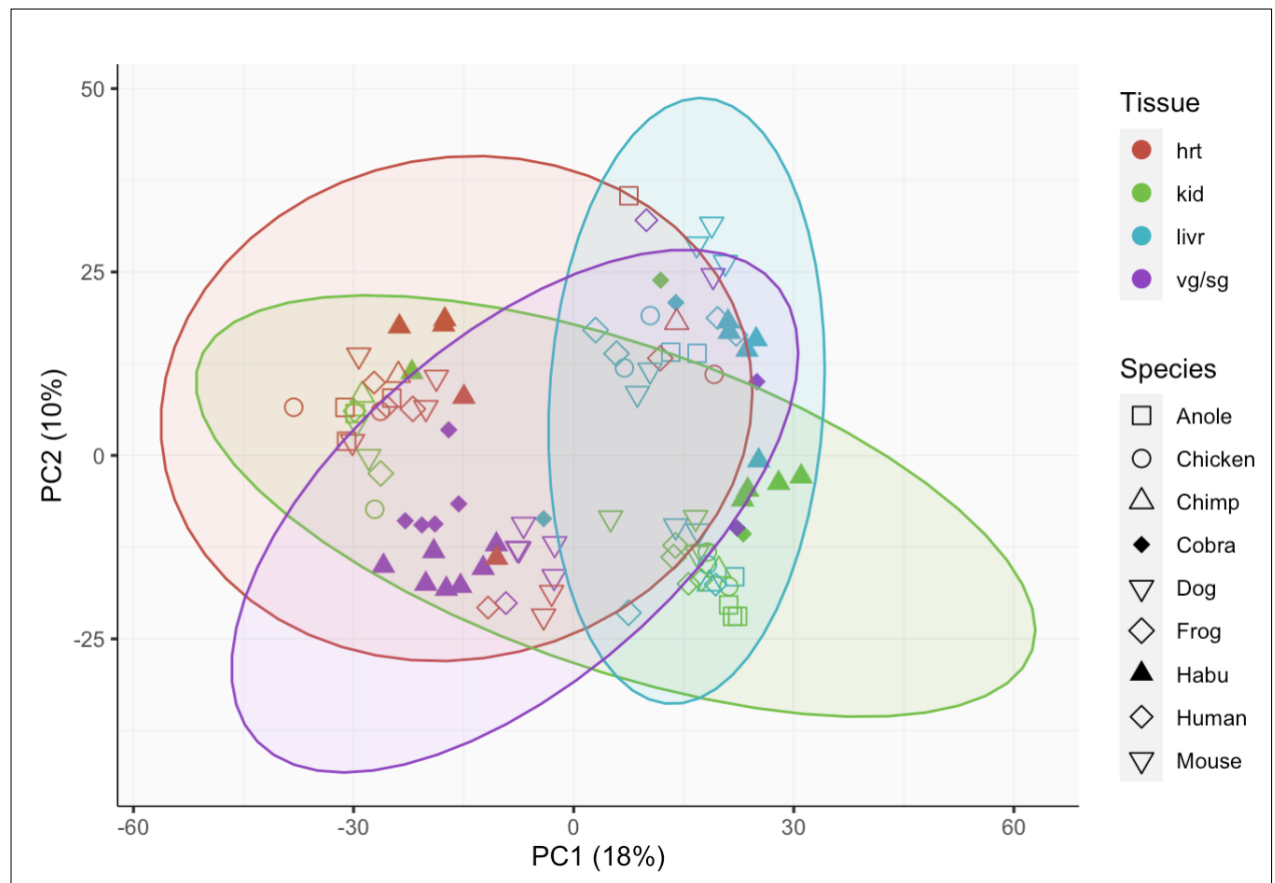
Figure S5: As a sanity check we chose orthologs at random to check whether the transcriptomes would still be clustered by tissue; however, a random set of genes produced no clustering. Selection of random gene sets was done using a random number generator in R (Online supplementary material)
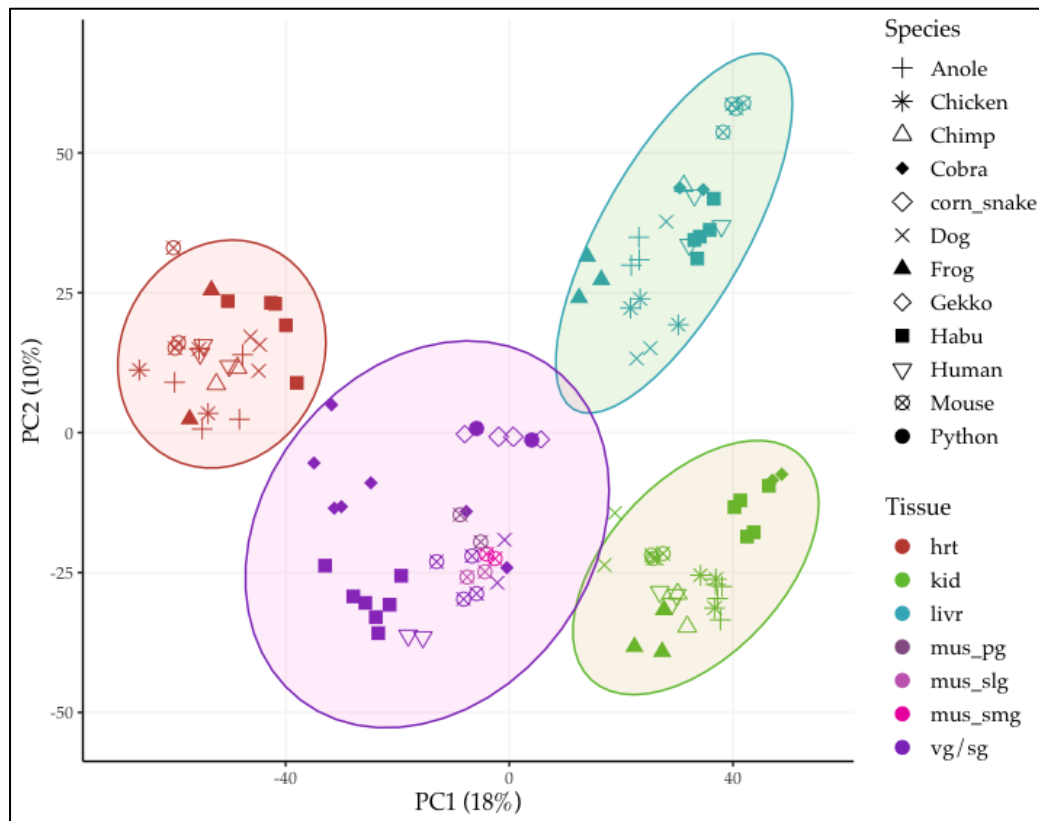
Figure S6: PCA using all 2291 expressed orthologs between 12 taxa. This PCA contains additional data salivary glands of non-venomous reptiles like royal python, corn snake, and leopard gecko. In addition it also contains transcriptomes from different morphologies of mouse salivary glands. Even with this reduced dataset, venom and salivary glands still cluster together.
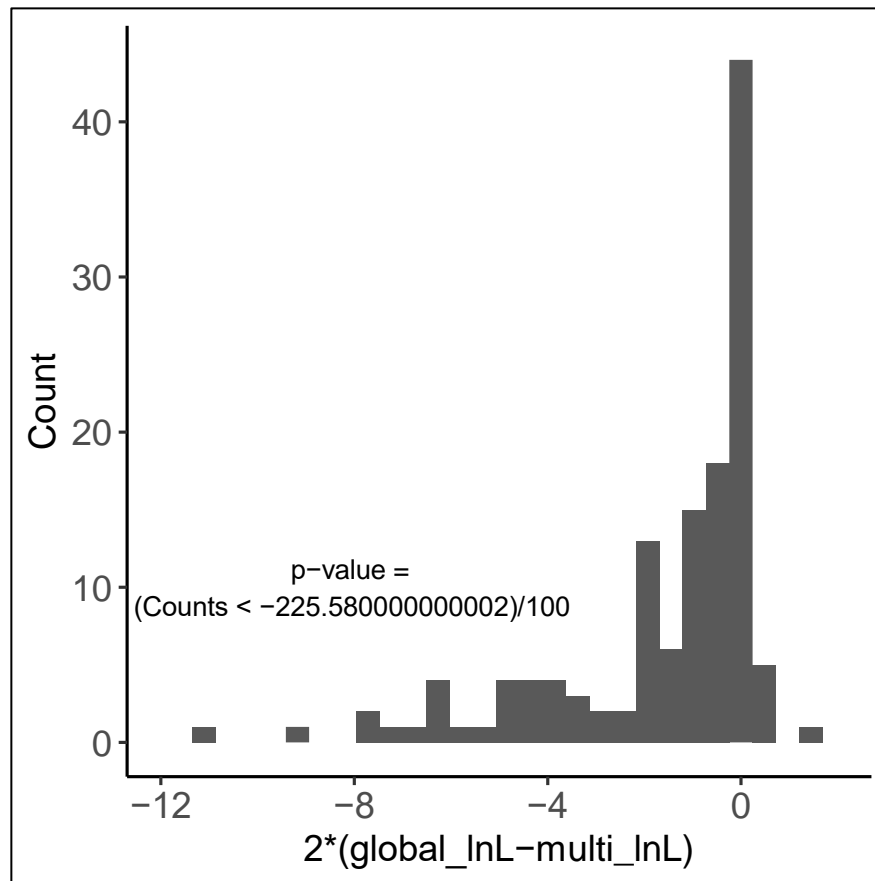
Supplementary Figure M5: Null distribution obtained from simulations using lhtest.R script provided by CAFE. Significance is determined by how far left the observed likelihood ratio (2 x lnLglobal - lnLmulti) would fall on the tail of the distribution. In our case the likelihood ratio count would fall on the far left of the distribution indicating a very low *p* value.